

Title: Preliminary proposal to add a new provisional kIDS property (Unihan)

Author: Ken Lunde & John H. Jenkins

Date: 2021-06-30

The purpose of this document, which is a preliminary proposal to add a new provisional Unihan database property, *kIDS*, is threefold:

1. Outline the standardization timeline
2. Identify any barriers early on in the process through constructive and meaningful feedback from both the UTC and IRG
3. Solicit help in collecting ideograph components for use in IDSes

IDS is an abbreviation for *Ideographic Description Sequences*, which is documented in [Section 18.2](#), *Ideographic Description Characters*, of the Core Specification.

The standardization timeline has two targets: Unicode Version 15.0 (2022) and Unicode Version 16.0 (2023).

Unicode Version 15.0 (2022)

The targets for Unicode Version 15.0 are to:

1. Encode four or five new Ideographic Description Characters (IDCs)
2. Encode a modest number of ideograph components as new CJK Unified Ideographs for use in IDSes

This then sets the stage for adding the provisional *kIDS* property in the subsequent version of the Unicode Standard.

New Ideographic Description Characters

Four new IDCs were most recently proposed in [L2/18-012](#) (aka [IRG N2273](#)) as shown in the table below:

IDC	Type	Character Name
	Binary	IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM RIGHT
	Binary	IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM LOWER RIGHT
	Unary	IDEOGRAPHIC DESCRIPTION CHARACTER HORIZONTAL REFLECTION
	Unary	IDEOGRAPHIC DESCRIPTION CHARACTER ONE HUNDRED EIGHTY DEGREE ROTATION

The [Ideographic Description Characters](#) block, which is the most appropriate block for encoding these four new IDCs, has exactly four available code points: U+2FFC through U+2FFF. The proposal will recommend encoding these new IDCs using these particular code points.

There is a fifth new IDC that is being considered whose recommended name is IDEOGRAPHIC DESCRIPTION CHARACTER STROKE SUBTRACTION. This new IDC is binary, and is followed by two components:

1. An ideograph component
2. A CJK stroke from the [CJK Strokes](#) block that is omitted from the ideograph component, or multiple CJK strokes if an IDC is used

Below are examples that use U+2296 \ominus CIRCLED MINUS as a placeholder for the actual IDC:

- The IDSes for U+2002A 𠄎 and U+2002B 𠄏 are difficult to represent with existing ideograph components, but could be represented as \ominus 其 \swarrow and \ominus 其 \searrow , respectively.
- The IDS for U+2CEBB 豸 is also difficult to represent with existing ideograph components, but could be represented as \ominus 豕 $\begin{smallmatrix} \square \\ \square \end{smallmatrix}$ $\swarrow \searrow$.

A counter-example for the first example above would be to instead encode the common ideograph component of U+5176 其, U+2002A 其, and U+2002B 其 as a new ideograph component, but that handles only this particular case.

New Ideograph Components

A non-trivial number of CJK Unified Ideographs include components that cannot easily be represented in IDSeS, and a modest number of additional ideograph components—encoded as new CJK Unified Ideographs—would serve to improve the IDSeS of such ideographs.

As a mere sample, the table below illustrates the numbered ideograph components based on the 2021-05-26 version of [IDS.TXT](#), and provides examples of ideographs whose IDses include them (ideographs for which there is not an available font are shown as a placeholder glyph followed by their IDses shown in green and enclosed in red brackets):

Syntax	Component	Examples
{01}	𠂇	廌焉鳥勛翠埜忽憊旃鴻崙罵烏鴉鵠襦襦迥鞫
{02}	𠂆	免晚象劓劓勛龔為終鯁渙鯁芻鬻毚
{03}	𠂅	夤奐媿夤劬夤舛憊揜揜寔槩椀為煖穈羹蠮象貌輓鵠 筭龔瘕
{04}	𠂄	蒙蒙
{05}	目	段瞢段段鋤段儺段段儼
{06}	丰	段矐段瞢緞菽逡鰕

sion block that includes sufficient unused code points for encoding future ideograph components should be considered. One suggestion is to use U+2FA20 through U+2FAFF for this new block, which provides 224 code points.

Unicode Version 16.0 (2023)

With the four—or five—new IDCs and ideograph components encoded in Unicode Version 15.0, the target for Unicode Version 16.0 is to add the provisional *kIDS* property to the UniHan database, based on [IDS.TXT](#). This IDS database, which is developed and maintained by Andrew West, has no copyright and is not encumbered by a license.

Delaying the addition of this new provisional property until Unicode Version 16.0 gives sufficient time for this IDS database to incorporate the new IDCs and new ideograph components, and also to make continued refinements to other IDSes. This should result in more robust property data.

That is all.